



RESEARCH REPOSITORY

*This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination.
The definitive version is available at:*

<http://dx.doi.org/10.1007/s10758-015-9249-5>

Gibson, D. and de Freitas, S. (2016) Exploratory analysis in learning analytics. Technology, Knowledge and Learning, 21 (1). pp. 5-19.

<http://researchrepository.murdoch.edu.au/id/eprint/25685/>

Copyright: © Springer Science+Business Media Dordrecht 2015.
It is posted here for your personal use. No further distribution is permitted.

Exploratory Analysis in Learning Analytics

David Gibson¹ and Sara de Freitas²

1.Curtin UniversityBentley, PerthAustralia

2.Murdoch UniversityPerthAustralia

Abstract

This article summarizes the methods, observations, challenges and implications for exploratory analysis drawn from two learning analytics research projects. The cases include an analysis of a games-based virtual performance assessment and an analysis of data from 52,000 students over a 5-year period at a large Australian university. The complex datasets were analyzed and iteratively modeled with a variety of computationally intensive methods to provide the most effective outcomes for learning assessment, performance management and learner tracking. The article presents the research contexts, the tools and methods used in the exploratory phases of analysis, the major findings and the implications for learning analytics research methods.

Keywords: Learning analytics; Computationally intensive mixed methods research; Game-based learning; Virtual performance assessment

1 Introduction

The promise of learning analytics as a new field of study in education and learning sciences has been highlighted in early studies (Shum and Ferguson 2012; Siemens 2012). However, while several studies have revealed the importance of using data, few have highlighted which analytic methodologies are needed at what points in the cycle of research to improve the efficacy of higher education teaching and learning practices.

Alongside this methodological challenge, analytics challenges have dramatically increased since new digital user interfaces and web based search approaches have evolved in recent times, leading to a wider expectation of personalized feedback. The increased complexity in the capabilities and roles presented in interactive learning experiences and a wider capacity to collect rich learner data have not always been used to inform feedback or to improve the learner's experience to date (de Freitas 2014; Quellmalz et al. 2012). Despite these challenges the infusion of data science methods and techniques into learning and behavioral science research are providing new tools for giving students more feedback when needed and allowing broader personalization approaches to be adapted for example through learning management systems (Gibson 2012). For these innovations to be adapted more widely into learning and teaching practices, new quantitative methods as well as a reconceptualization of mixed methods (Tashakkori and Teddlie 2003) need to be adopted.

This paper aims to discuss applicable exploratory methodologies for learning analytics that utilize 'computationally intensive' mixed methods in the early stages of research. We use the phrase 'computationally intensive mixed methods' to distinguish from the broad term 'mixed methods' that combines qualitative and quantitative methods (Creswell 2003). We specifically mean a new mixture of data mining and visualization mixed with a range of regression methods; and within the regression methods, a mixture of linear, nonlinear and symbolic methods. Roughly speaking, this mixture is a search for best fitting lines, curves as well as classes of equations to represent relationships in the data and build models of the phenomena under study.

The article summarizes our observations and the implications for research methods that we have garnered from two learning analytics research projects. One project is connected to work at Harvard University, and one is an ongoing retention analytics project at Curtin University.

The cases share some common features. First, the case studies involve *high-resolution data collection*. High-resolution data in this context means two or more orders of magnitude in the number of records per subject compared to the norm in educational research, from tens of records to hundreds or thousands of records. Compared to the previous norms in educational research in which a pre-test and post-test might be two of only a handful of data points, these large datasets offer new ways to consider data that should be exploited both in terms of closing the loop with users and the need to explore and model longitudinal information about learner interactions and performances. This longer duration and more highly diverse and detailed data for understanding the learner implies a capability to provide more comprehensive information to them and their tutors, more scaffolding and feedback for an end-to-end learning experience.

Second, the cases also illustrate the role of *technology as an interactive agent* in the production of data, because some of the data arises from a complex interaction product between the learner and the digital learning environment as well as from co-production of data by the learner, environment and social context. These co-production situations produce highly variable and diverse data sources that must be reconciled and interwoven. In particular, we assert that due to this feature, new psychometric challenges are emerging that cannot be dealt with by standard methods. Third, the data in each of the example cases needs to be *analyzed relatively quickly* in order for people to react, or the learning environment to adapt to the learner.

These three features of volume, variety and velocity help explain the rise in interest in data science methods in higher education learning analytics. The methods described here are associated with ‘big data,’ where information is flowing rapidly, is highly varied in format and grain size, and where the unit of analysis is quite small compared to the aggregated level where decisions need to be made (IBM 2015)

Finally, when the aim is to better understand how people learn or how and to what degree they have acquired knowledge, there is a need for exploratory and iterative psychometric modeling appropriate for each situation. Since the well-known psychometric models in traditional tests do not readily apply in highly interactive digital learning environments, there is considerable uncertainty about the affordances of the performance spaces, how people will leave traces through the space, and what forms of internal and external validity will apply to the meaning of the traces (Choi et al. 2010; Clarke-Midura et al. 2012; Eseryel et al. 2013). It seems natural then to assume that an iterative exploratory stance is called for when analyzing a newly constructed and little-studied virtual performance space. We will first outline the theoretical foundations of our research and then share the case examples in order to discuss key exploratory analysis methods and findings.

2 Data Science and Exploratory Psychometrics

Data science methods have arisen to deal with quickly accumulating, highly diverse data at high volumes such as the signals from atom-smashing experiments, large-scale distributed sensor network data, real-time traffic flow data and determination of neural pathways during learning and performance. Learning analytics in certain circumstances approaches these same levels of complexity, for example during the analysis of data from a brain-computer interface (Gonzalez-Sanchez et al. 2013), or an analysis of the recent history of a university's student retention data (Deloitte 2010). The cases we will share here are such examples. In these kinds of cases, learning analytics needs to integrate new data science methods into the traditional panoply of educational research methods and provide mechanisms for analyzing digital learning experiences with *high resolution and time sensitive data involving a diversity of issues* such as the role and impacts of social and human-to-machine communications, automated learning assistants (Hasler et al. 2013; Nair et al. 2004), learner decisions and judgments by learners as well as experts (Eseryel et al. 2013), characteristics of multidimensional items and constructs (Behrens et al. 2011), effects of various media as performance prompts and the

affordances of digital practice and performance workspaces (Connolly et al. 2012; Dunleavy et al. 2008).

A technology-enhanced affordance for learning, for example, might include a digital pedagogical or conversational agent (Morris 2002; Sabourin et al. 2011) that is tutoring or giving hints to someone during a learning experience. Analytics might aspire to find patterns in a learner's decisions during a digital game or simulation (Christensen et al. 2011; Clarke and Dede 2010; Rupp et al. 2010). An assessment process giving rise to data for analytics might include students reviewing and commenting on each others' digital creations through open-ended online discussions (Ertmer et al. 2007; Van Der Pol et al. 2008; Webb 2010) and involving texts that are increasingly amenable to natural language processing (Jordan 2009). A technology-enhanced summative assessment might include a multimedia constructed-response item created with an online animation and modeling application (Lenhard et al. 2007; Mislevy 2011). In addition, new targets for learning made possible by technology-enhancements might be measured on a student's collaboration capabilities based on their responses to receiving remote asynchronous expert feedback about how they worked with each other to solve a problem and communicate their understandings (Rissanen et al. 2008). Or their emotions might be monitored as they work in an engaging virtual world experience that unobtrusively documents progression of their leadership and ethical development over time (Turkay and Tirthali 2010). This wide range of potential targets of an analysis highlights the need for additional qualifiers whenever the term 'learning analytics' is used, in order to ensure that appropriate theoretical and methodological frameworks have been selected to guide the research process. Here, we are asserting that reporting on the exploratory modes in both analytics and psychometrics helps clarify a study.

Exploratory analysis is concerned with understanding the structure and relationships of data (Morgenthaler 2009; Tukey 1977) which can lead to new questions and hypotheses. It can be thought of as an earlier step in the research process than hypothesis formation; however, we hold that the approach should be part of an iterative process of any well-constructed computationally intensive mixed methods research program. Psychometrics is the science of measuring mental capacities and

processes (Kline 1998). We hold that new psychometric models are needed in cases where the subject is interacting in a dynamic digital environment.

We assert that the research process must utilize iterative exploratory stances when the digital spaces are relatively new and unexplored. A synthesis of methods, perhaps unique to each digital performance space and using computationally intensive mixed methods, we believe is helpful to lead to observations and inferences about what someone knows and can do *based on their evolving interactions in a digital media learning environment*. These observations are provisional during the exploratory phase of research and are not fully codified due to several features of interactive digital learning environments (e.g. time and space dilation, dynamic interactions with the computer, feedback cycles, complexity and entanglement), which make the streams of data more complex than snapshots of a process frozen in time (Gibson and Jakl 2013).

3 Two Cases of Learning Analytics in Higher Education

The first case concerns the analysis of learner actions in a game-based virtual performance assessment created at Harvard (Clarke-Midura et al. 2012) and analyzed by one of the authors (Gibson and Clarke-Midura 2013). The second case concerns a longitudinal retention study based on 5 years of performance data from 52,000 students from a wide range of distributed sources of information (de Freitas et al. 2014). These two cases highlight key issues that arise when undertaking learning analytics studies in higher education and point to a need for a new form of mixed methods when working with big data in learning analytics research.

3.1 Virtual Performance Assessment

The first of the two learning analytics cases concerns the question of whether and in what ways user actions in a game can be predictive of the final assessment grade related to knowledge and skills acquisition. The case comes from a purpose-built game of the Virtual Assessment Project at the Harvard Graduate School of Education (Gibson and Clarke-Midura 2013), which has an educative

aim to examine whether the game is able to assess middle school students' abilities to design a scientific investigation and construct a causal explanation (Clarke-Midura et al. 2010). The assessments start out with one of two problems that students must solve: Why is there a frog with six legs? What is causing a population of bees to die? The assessments were designed in the Unity game engine (<http://unity3d.com/>) and have the look and feel of a videogame (Fig. 1).

Participant actions (e.g. opening a page, saving a note) were time-stamped and labeled as an event. Analytic data from two pilots consisted of 1987 users (423,616 event records) in the frog assessment and 1958 users (396,863 event records) in the bee assessment. The analysis included demographic information about students (age, gender, class, teacher), the starting prediction for the cause of the problem, raw event data (e.g. up to when a student made their final claim about the problem) and human-scored constructs of designing a causal explanation and designing a scientific investigation.

The purpose of the analysis was to search for *patterns of action* that might relate to the performance of the user *correlated with the student's final claim*. Could the action log and score data tell us about the user's performance? Additional questions included:

- Is there a relationship between overall duration and score level?
- Were there performance differences that differed by gender, age, and grade?
- Was there a relationship between someone's prediction at the beginning of the assessment and their claim at the end?

To address the question about the relationship between a student's prediction and claim, we constructed a matrix of the *empirical probabilities* (the relative frequency of an outcome in relation to a number of trials) constructed from a count of the population's actions which assist in creating Bayesian inferences about the scientific problem-solving path-maps of learners (Stevens et al. 2005). The probabilities can also be constructed for a subset or a comparable population, and can be built with a focus on any action or set of actions. In this case, we chose to count the prediction and claim actions by the whole population. In the bees and frogs assessments, each student made a prediction, then explored the digital space collecting evidence that was used to make a claim at the end of the

session. Our matrix thus had seven categories of ‘prediction’ rows (e.g. p1, p2, ..., p7) crossed with 5 categories of ‘claim’ columns (e.g. c1, c2, ..., c5). The empirical probability of a student making the prediction ‘p1’ is the row sum for p1 over the sum of all predictions. The probability of a shift in thinking from prediction to claim is the conditional probability of predicting some claim given both the probability of claiming and the prior probability of the prediction, using standard probability theory.

We concluded that in a similar population of middle school students, we would expect that 38 % would eventually make the same claim as an expert but only 8 % would begin the assessment with that prediction. So, since a significant portion of the test takers arrived at this conclusion after interactions in the virtual assessment, this indicated to us that *the virtual performance assessment experience might itself be educative* and that *user actions might provide evidence of learning as well as thought processes*. This early result could also be obtained by traditional pre- and post- methods in a multivariate approach; however, the point here is that the analysis at this stage is viewing the landscape of potential ending points (the claims) as probabilities of trajectories in the space of possibilities of action rather than as factors in a result. The ‘pre’ condition of the choice of a prediction by the student is a starting point for trajectories that help explain the effect of working in the space with certain materials over time, and in some order, to lead the student to a supported conclusion. The analysis we are seeking has more to do with the journey than the destination, and as much to do with the clusters and sequences as the final resting position on the ‘right claim.’

Before we explored more specific action patterns, such as which doors did people open and who did people talk to, we wondered how the total amount of time spent in the simulation would relate to the outcome. Symbolic regression using the free educator version of Eureqa (Schmidt and Lipson 2009) was selected as the method to obtain a specific predictive mathematical expression for the relationship between overall duration and score level. The mathematical expression goes beyond the correlational idea of ‘more time = higher score’ to provide a predictive equation of the optimal time and configuration for highest performance. The symbolic regression method produces a range of equation options representing a class of equations that best meets a fitness criterion, which serves as the goal of

an evolutionary algorithm search. In all cases reported here, the method included cross-validation of 33 % of the data set for training to 66 % of the data used for prediction, in randomly selected tests conducted thousands of times for each proposed solution.

The analyst then selects a solution from the range of options resting on the Pareto Curve representing the trade-off in efficiency between error and complexity: the less complex the mathematical expression, the higher the error and vice versa (Fig. 2). That selection is then one equation in a constructed set of nonlinear partial differential equations that describe the network of causative relations in the event window, so the choice of the solution on the Pareto Curve is a *modeling decision* with implications for the overall complexity and semantics of the final complete set of algorithms of the network. This example provides evidence that *predictive (generally nonlinear) relationships can be discovered via symbolic regression*.

To explore action patterns more deeply, we then conducted rule discovery using the ‘Apriori’ algorithm (Hegland 2005; Witten and Frank 2005) to find the most prevalent co-occurrences of actions. We used this and other algorithms in the data mining toolset ‘Weka’ (Hall et al. 2009) to perform exhaustive searches and optimization routines, which resulted in a *descriptive and associative rule set* (compared to the *deterministic or causative mathematical rule set* of the symbolic regression method). Such an associative rule set, when considered with the confidence of a rule discovered by symbolic regression, can elucidate the hierarchical (spatial) as well as temporal structure (Campanharo et al. 2011) of the relationships in a virtual performance assessment created by the paths of multiple users traversing the space and utilizing resources.

The discovery of association rules among qualitative data can also lead to network representations that are amenable to additional analysis (Han et al. 2007). A digraph (Fig. 3) is such a network representation which allows visual inspection comparing it to alternative graphs made by less successful students (not shown). A digraph was created based on the *association rule network* (Baker 2010; Han et al. 2007) for each of the cells in the empirical probability matrix. The important link here is to the time and space complexity of the digital learning environment, which underscores

the need for iterative exploratory analytics as well as iterative psychometric inferences in order to begin to understand what groups of students have done in the virtual space.

In a digraph the edge from one node to another has a directional meaning—as in causality or implication. Thus, an association rule can be directional; for example if the rule says that 100 % of the people who conduct ‘research_3’ then go on to conduct ‘research_1’ but 0 % conducted those activities in the reverse order, we can draw a circle for research_3 with an arrow leading to research_1 (Fig. 3). We found by visual inspection of the digraphs for each pair of predictions-to-claims in the empirical probability matrix, that students who did not have a structure of scientific investigation similar to Fig. 3 were more likely to have missed important evidence and reached a weaker conclusion.

We also used a clustering method on event-level user actions based on closeness determined by the ‘Expectation Maximization’ algorithm (Dempster et al. 1977) and found that clusters of events mapped closely to claims, but were more complex, because they were formed from all available data. For example, students who shared similar search and resource utilization strategies might be clustered together, even though they reached different conclusions about the data and made different claims. We found that *action pattern sequences* could be identified only with the aid of the virtual assessment designers (domain experts) and that these higher aggregations of action-sequences that we called ‘*motifs*’ were more meaningful units of analysis than without the domain expert knowledge of the virtual performance space. In this sense, the data cannot speak to the intentions of the designers without the higher-level knowledge of meaningful sequences of actions. For example, it was not informative to cluster all the ‘open door’ actions, but instead to cluster the doors with geographic knowledge of the virtual space (e.g. was it a door to a particular lab or farmhouse, where one could have a conversation with a particular actor agent).

In this case example, analysts were able to predict one of the claim groups based on visual inspection of the absence of resource utilization. The exploration by clustering methods also led to the analysts making structural recommendations to change the game design as well as the data acquisition labeling, without which all students had almost identical access to all the affordances of the virtual

performance assessment, rendering cluster analysis impotent for learning analytics. An alternative game design model could channel a participant into different subsets of affordances based on previous decisions. However, a channeled structure tends to force a limited set of choices and thereby restrict the user's degrees of freedom, so care has to be taken in the design of both the learning experience and data labeling for collection.

3.2 Retention Analytics

In the second case, the analytics team aimed to understand why some students succeed in higher education courses of study while others drop out, so that appropriate support and academic guidance can be provided. An initial study was undertaken in 2010 in the Curtin Business School and a second larger study across the entire university was recently concluded in 2013, both with significant technical and research assistance from Deloitte Consulting Services.

Our main findings supported two hypotheses on student retention in the wider literature, and confirmed the importance of academic and curriculum engagement for supporting students in the first year of undergraduate study. For example, research-based observations have been made concerning higher rates of retention for international than for Australian students (Olsen 2007) and our results tended to confirm these findings. However, within the category of international students, our new analysis pointed to a need to understand the specifics of who is retained and who drops out, how students are similar to and differ from others, how behavioral patterns and retention results vary across the university's offerings, and what can be done to intervene and improve retention, since the cost of attrition is estimated to be about 2 million dollars for every 50 students who leave in the second year (de Freitas et al. 2014).

The analytics methodology in this case followed *a staged process of data acquisition, preparation, discovery and analysis*, which centered on the creation of a self-organizing map (SOM) that was further shaped by the analytics team (Fig. 4). The SOM is a *semi-supervised machine learning method* based on a *multidimensional similarity metric* that groups together related data clusters of individuals who share similar behavioral features, or attributes. It is semi-supervised because the

analytics team intervenes in cycles of machine learning to feed in newly transformed data into the modeling process as issues such as collinear relationships (data that reinforces and overstates its impacts on a model) are discovered.

The staged data collection and analysis process engaged a wide spectrum of key stakeholders, with data sources and hypotheses statements coming from and going back into a series of public engagements across the university. The study thus used robust traditional mixed methods research with both quantitative and qualitative foundations for data collection and analysis, and then during the analysis phases, additional *computationally intensive mixed methods* as defined in this article.

Data sources included several large sets of anonymous, post hoc data from the university's student data application, the learning management system, post-unit student evaluation surveys, online library, interviews and focus group sessions, concentrating on a four-year period of the university's recent history. Since some students exit and then re-enter the university after more than 1 year has elapsed, the data preparation phase included *time-event segmentation* in the process of defining the data that was *in scope*. (This method differs from standard year-on-year retention accounting in what we termed 'lifetime student retention'). The analytics team *linked additional external datasets* including census data, social economic status indexes and geocoding.

Over 1200 attributes for each student were considered (62 million elements), which led to a need for *data reduction* methods such as not including a set of data that was known to be the opposite of another set (e.g. males vs. females) and sets of variables with 1:1 relationships signifying in many cases, multi-collinearity (e.g. the combined effect of linked variables that lead to overstating their influence and drowning out the influence and details of other variables). A 13 million-element analytic data set (ADS) was extracted and served as the basis for the SOM *semi-supervised machine learning model*, which was trained on 52,000 students with over 250 attributes, including some online students and all part-time and onsite students at the Bentley main campus. The ADS dataset can now support future cross-validating analyses, for example, using statistical and symbolic regression, neural networks and social network analysis as well as alternative data mining and pattern discovery

methods, to be conducted by a Teaching and Learning research group that has been established to build the university's analytics capacity.

During a four-week data collection period, while all available data sources were being combined, cleaned, and made ready for data exploration, the analytics team held a series of workshops and focus groups with staff and students to create a wide range of hypotheses about the possible causes and conditions of retention and attrition. The focus group sessions were designed to provide hypotheses to be used to query the ADS. Over 200 hypotheses were generated by these 'diverge workshops' and were then tested by the analytics team against the best-fit SOM model of the data. A second workshop series with stakeholders then tested 50 of the key hypotheses for plausibility with the model using visual inspection, and involved stakeholders in interactive group exploration of the data made possible via interactive visualizations created with support and evidence from the data's statistical characteristics. The analysis of the dataset has been made available in (de Freitas et al. 2014)

Since the initial analysis, which uncovered hidden clusters of students at risk who would not have been identified prior to the research, the student services team at Curtin University has further explored interventions and is re-thinking its strategies and planning for future data refreshes on a daily basis. The project kick-started two learning analytics working groups, one in the Information Technologies group devoted to building an enterprise data bus to integrate the university's information systems and one in Teaching and Learning to build the university's research capacity to utilize a continuous stream of student information.

4 Discussion

The case studies have both shown how differently data can be used to understand more completely the learner's experience. In the game example we could see how the individual's experience can be modelled and mapped effectively and feedback can be provided throughout the experience. In the retention study the benefit of using performance data throughout the life cycle of the learner shows how interventions can be better planned and deployed at the right moments during the student's study.

Both cases illustrate the potential of performance driven metrics, but also show the value of using more than one method of analysis.

The cases outlined are emblematic of learning analytics research on digital learning experiences that require exploratory data analysis (Tukey 1977), that is, an analysis which does not start with a hypothesis, but searches initially for patterns in the data in order to discover broad sets of questions and potential hypotheses that require further study.

Visual inspection was an important step in the exploratory process in the cases. In addition, determining features for exploratory analyses are, among other things, the volume, variety and velocity of the data. The records-per-unit-of-analysis described here for example, varied across the three cases from 175 to 250 per subject in while the total data-element sizes of the data based varied from 31,000 to 13 million.

Data mining methods (e.g. clustering, machine learning, symbolic regression, network analysis) were used in the cases; additionally, causal explanations (e.g. using empirical probabilities, network analysis and symbolic regression) were sought in the cases, driven by the research questions. A number of methods and tools new to our team were applied in the course of these studies, which has begun to build a set of methodological ideas and approaches for application in future research. We have highlighted these implications in the summaries below.

Transforming data for data mining in all cases involved both reduction moves and intermediate pattern aggregations which do not reduce the data. For example, the clustering in Case 1 was ineffective until subject domain experts identified a two or three-element chain of actions we have called a motif. In Case 2, reductions followed traditional lines (e.g. finding multicollinearities), but in addition, the SOM is also a form of motif creation via transformation since the unsupervised and supervised phases of the semi-supervised machine learning method chooses a different subset of attributes for each cell of the map to maximize the local similarity of neighboring groups.

The advantage of unsupervised machine learning in this case was that no prior hypotheses or assumptions biased the model, and the resulting best-fit map is then available for multiple hypothesis testing as overlays on the SOM, while the supervised reduction of the dataset allowed finer and finer grain sizes of the underlying population similarities. This method allowed us to test quite a large number of hypotheses compared to traditional single hypothesis testing research methods.

The first case, the Harvard virtual performance assessment in science, illustrates several features of exploratory data analysis of digital learning experiences. First, the *context was captured in the event data along with the learner action, decision, and product*. However, that context needed to be effectively reconstructed from the smallest items of data into larger clusters of information. For example, a data element named ‘opened door’ by itself was relatively meaningless compared to knowing that it was a particular door, opened after another significant event such as talking to a scientist. Thus, *patterns of action* were transformed into *motifs*, which then became the *transformed units of analysis*. This concept of the unit of analysis containing the context for interpretation may be a new methodological requirement for the analysis of digital learning experiences and needs further study. It also highlights the fact that once the context is understood, the motifs themselves can become features for machine learning training, as well as for embedding in future iterations of the digital learning experience as filters or triggers for immediate feedback to the learner. These features may not be identified or understood until a few iterations of the exploratory process have occurred.

Second, as a large number of users traverse through the network of possibilities in a digital problem space, key movements of the population within the network can be counted and then used as the basis for *empirical prior probabilities*. In particular, each pathway in such a network can be further characterized or specified with a predictive nonlinear mathematical relationship found for example through *symbolic regression*. Or, alternatively an *association rule network* can be created that distinguishes user action patterns and motifs by the count prevalence of the trajectories from one resource to another. For example, if 100 % of the population goes to resource 3 after resource 1 (skipping over and not utilizing resource 2), then with a very high probability (and if the sample is a good sample of the greater population), the next user entering the system will follow that path and a

highly probable inference can be made about what the person now using resource 1 will do next. A limitation of the research reported here is that we were not able to test whether the probabilities were a good basis for prediction in a future development cycle of the science game. This particular finding is thus limited by being a post hoc analysis, and needs to be validated by a future iteration of the science game.

Third, the complex set of data relationships discovered by the exploratory methods in case 1 bore *a meaningful structural relationship to the problem space*; this is possibly a generalizable observation, but needs to be studied further. Validating evidence for generalizability in other studies can be found. For example in structural studies of the brain based on the networks entailed during certain tasks and that differ from task to task (Sporns 2011). Additionally, a *cluster analyses* of a problem space can reveal that some resources are critical to success and others are ignored and not important to the most successful learners (Quellmalz et al. 2012) or a *network visualization* of the problem space can highlight how people relate to each other and a task such as quoting and using scientific resources (Bollen et al. 2009).

In the second case, exploratory methods assisted in *a staged process of stakeholder involvement in data acquisition, preparation, discovery and analysis*. A *semi-supervised machine learning* method used during the exploratory phase utilized an unsupervised *SOM* that was created based on a multidimensional similarity metric and facilitated the *simultaneous testing* of over 50 hypotheses. The three aspects of big data—volume, variety and velocity—are most noticeable in this case. The volume of data comprised 13 million records for 52,000 subjects, or about 250 records per subject. A wide variety of data was collected from ten digital sources that included study patterns, performance in units of study, attendance, survey question responses, demographic profiles, library records, and other diverse sources. The velocity element was represented by the short timeline for data acquisition, preparation, discovery and analysis managed in stages and engaging over 200 people in focus groups and feedback sessions. Here as in the other cases, *nonlinear and symbolic mixed methods* were combined iteratively with *linear methods* to obtain results, perhaps defining a specific role for exploratory data analysis in *computationally intensive mixed method research*.

5 Conclusion

The paper has revealed how powerful learning analytics can be when computationally intensive mixed methods are used to ensure that multiple points of data are integrated during a study. Methodological observations from the two learning analytics research projects illustrates the wide scope of data set sizes and resolutions and the new analytic methods that are now becoming increasingly available for research in the learning sciences. The computationally intensive mixed methods briefly outlined in this article are part of an expanding framework of educational research methods that enables researchers to deal with complexity in time and event structures involving complex data in higher education digital learning experiences. These methods are part of a new basis for iterative exploratory analyses and psychometrics, which is emerging as researchers undertake to understand what learners know and can do based on their interactions in digital learning spaces.

This paper brought together lessons learnt from two studies using large data to improve learning and teaching and provided our insights concerning methodology and tool selection. We argue in favor of mixed approaches that include computationally intensive exploratory stages that iterate between analytic findings and psychometric inferences. As these methods become more automatic they will support more timely and effective feedback to students and tutors through in situ co-construction of performance tracking as well as monitoring information about learning and teaching.

Future studies will aim to close the loop between the system and the learner by producing dashboards and feedback models that will allow us to better model learning patterns distilled from the interaction between students and digital learning environments. It is hoped that this work will advance the fields of learning analytics and psychometrics by for example, introducing fast feedback dashboards driven by analysis that has been designed alongside students and tutors.

While the study of learning analytics is relatively new, it is envisaged that there is a large potential for impact upon the learner's experience, through outreach and social networking, recommendations, more engaging learning experiences, rapid feedback, retention support and lifelong networking.

However, further study needs to be undertaken to explore how we can automate some of the methods

described here in order to provide effective feedback to the learner and tutor, what are the benefits and limitations of rapid feedback and how some of the more complex methods can be deployed to the best effect.

We predict that the fields of gamification, technology enhanced learning and learning analytics will have capacity to lever off one another to enhance the learning experience, and welcome deeper studies of large datasets requiring computationally intensive mixed methods that support new approaches in digital media teaching and learning.

References

- Baker, R. S. J. (2010). Data mining for education. *International Encyclopedia of Education*, 3, 112–118. doi:10.4018/978-1-59140-557-3.ch075
- Behrens, J., Mislevy, R., Dicerbo, K., & Levy, R. (2011). Evidence centered design for learning and assessment in the digital world. In M. Mayrath, J. Clarke-Midura, D. Robinson, & G. Schraw (Eds.), *Technology-based assessments for 21st century skills* (pp. 13–54). Charlotte, NC: Information Age Publishers
- Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M. A., & Balakireva, L. (2009). Clickstream data yields high-resolution maps of science. *PLoS One*, 4, e4803. doi:10.1371/journal.pone.0004803.
- Campanharo, A., Sirer, M., Malmgren, R., Ramos, F., & Amaral, L. (2011). Duality between time series and networks. *PLoS One*, 6(8), e23378. doi:10.1371/journal.pone.0023378.
- Choi, Y., Rupp, A., Gushta, M., & Sweet, S. (2010). Modeling learning trajectories with epistemic network analysis: An investigation of a novel analytic method for learning progressions in epistemic games. In *National council on measurement in education* (pp. 1–39).
- Christensen, R., Tyler-Wood, T., Knezek, G., & Gibson, D. (2011). SimSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*, 6(2), 201–220.
- Clarke, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research in Teacher Education*, 42(3), 309–328.
- Clarke-Midura, J., Code, J., Dede, C., Mayrath, M., & Zap, N. (2012). Thinking outside the bubble: Virtual performance assessments for measuring complex learning. *Technology-based assessments for 21st Century skills: Theoretical and practical implications from modern research* (pp. 125–148). Charlotte, NC: Information Age Publishers.
- Clarke-Midura, J., Mayrath, M., & Dede, C. (2010). Measuring inquiry: New methods, promises and challenges. *Library*, 2, 89–92.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers and Education*,. doi:10.1016/j.compedu.2012.03.004.

- Creswell, J. (2003). *Research design: Qualitative, quantitative and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.
- De Freitas, S. (2014). *Education in computer generated environments*. London, New York: Routledge.
- De Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., Dunwell, I., Arnab, S. (2014). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*. doi:10.1111/bjet.12212.
- Deloitte. (2010). *Student retention analytics in the Curtin business school*. WA: Bentley.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, 39(1), 1–38.
- Dunleavy, M., Dede, C., & Mitchell, R. (2008). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, 18(1), 7–22. doi:10.1007/s10956-008-9119-1.
- Ertmer, P. A., Richardson, J. C., Belland, B., Camin, D., Connolly, P., Coulthard, G., et al. (2007). Using peer feedback to enhance the quality of student online postings.... *Journal of Computer-Mediated Communication*, 12(2), 1–15.
- Eseryel, D., Ifenthaler, D., & Ge, X. (2013). Validation study of a method for assessing complex ill-structured problem solving by using causal representations. *Educational Technology Research and Development*, 61, 443–463. doi:10.1007/s11423-013-9297-2.
- Gibson, D. (2012). Game changers for transforming learning environments. In F. Miller (Ed.), *Transforming learning environments: Strategies to shape the next generation (advances in educational administration)* (Vol. 16, pp. 215–235). Bradford: Emerald Group Publishing Ltd. doi:10.1108/S1479-3660(2012)0000016014.
- Gibson, D., & Clarke-Midura, J. (2013). Some psychometric and design implications of game-based learning analytics. In D. Ifenthaler, J. Spector, P. Isaias, & D. Sampson (Eds.), *E-learning systems, environments and approaches: Theory and implementation*. London: Springer.
- Gibson, D., & Jakl, P. (2013). *Data challenges of leveraging a simulation to assess learning*. West Lake Village, CA. http://www.curveshift.com/images/Gibson_Jakl_data_challenges.pdf
- Gonzalez-Sanchez, J., Chavez-Echeagaray, M. E., Lin, L., Baydogan, M., Christopherson, R., Gibson, D. et al. (2013). Affect recognition in learning scenarios: Matching facial- and BCI-based values. In *2013 IEEE 13th international conference on advanced learning technologies* (pp. 70–71). doi:10.1109/ICALT.2013.26.
- Hall, M., National, H., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11, 10–18. doi:10.1145/1656274.1656278.
- Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: Current status and future directions. *Data Mining and Knowledge Discovery*, 15(1), 55–86. doi:10.1007/s10618-006-0059-1.
- Hasler, B. S., Tuchman, P., & Friedman, D. (2013). Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Computers in Human Behavior*, 29, 1608–1616. doi:10.1016/j.chb.2013.01.004.
- Hegland, M. (2005). The Apriori algorithm: A tutorial. *Institute for Mathematical Sciences Preprint Series*. <http://www2.ims.nus.edu.sg/preprints/2005-29.pdf>
- IBM. (2015). Big data. <http://www-01.ibm.com/software/au/data/bigdata/>

- Jordan, S. (2009). Assessment for learning: pushing the boundaries of computer-based assessment. *Research in Higher Education*, 3(1), 11–19.
- Kline, P. (1998). *The new psychometrics: Science, psychology, and measurement*. London: Routledge.
- Lenhard, W., Baier, H., Hoffmann, J., & Schneider, W. (2007). Automatic scoring of constructed-response items with latent semantic analysis. *Diagnostica*, 53(3), 155–165. http://apps.isiknowledge.com.libproxy.unm.edu/full_record.do?product=WOS&colname=WOS&search_mode=RelatedRecords&qid=25&SID=2AhKOADF4oMFGj8O5P9&page=3&doc=30
- Mislevy, R. (2011). *Evidence-centered design for simulation-based assessment*. Los Angeles, CA: The National Center for Research on Evaluation, Standards, and Student Testing.
- Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*,. doi:10.1002/wics.2.
- Morris, T. W. (2002). Conversational agents for game-like virtual environments. In *AAAI 2002 spring symposium on artificial intelligence and interactive entertainment* (pp. 82–86). <http://www.google.com/url?sa=t&ct=res&cd=1&url=http://www.qrg.northwestern.edu/aigames.org/papers2002/TMorris02.pdf&ei=crZZR9yjNozssgKpgPHIBg&usg=AFQjCNHT8F92vjehn8upv4EK5JlZWP3Hrg&sig2=yBAN659b8qSvJ-IXns3pBQ>
- Nair, R., Tambe, M., Marsella, S., & Raines, T. (2004). Automated assistants for analyzing team behaviors. *Autonomous Agents and Multi-Agent Systems*, 8, 69–111. doi:10.1023/B:AGNT.0000009411.79208.f4.
- Olsen, P. (2007). *Staying the course: Retention and attrition in Australian universities findings*(pp. 1–16). Sydney. <http://www.spre.com.au/download/AUIDFRetentionResultsFindings.pdf>
- Quellmalz, E., Timms, M., Buckley, B., Davenport, J., Loveland, M., & Silberglitt, M. (2012). 21st Century dynamic assessment. In M. Mayrath, J. Clarke-Midura, & D. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 55–90). Charlotte, NC: Information Age Publishers.
- Rissanen, M. J., Kume, N., Kuroda, Y., Kuroda, T., Yoshimura, K., & Yoshihara, H. (2008). Asynchronous teaching of psychomotor skills through VR annotations: Evaluation in digital rectal examination. *Studies in Health Technology and Informatics*, 132, 411–416.
- Rupp, A., Gushta, M., Mislevy, R., & Shaffer, D. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment*, 8(4), 1–45.
- Sabourin, J., Mott, B., & Lester, J. (2011). Computational models of affect and empathy for pedagogical virtual agents. In *Standards in Emotion Modeling*. Leiden, NL: Lorentz Center International Center for workshops in the Sciences. <http://www.lorentzcenter.nl/lc/web/2011/464/presentations/Sabourin.pdf>
- Schmidt, M., & Lipson, H. (2009). Symbolic regression of implicit equations. *Genetic Programming Theory and Practice*, 7(Chap 5), 73–85.
- Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Educational Technology and Society*, 15, 3–26.
- Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. In *2nd international conference on learning analytics and knowledge* (pp. 4–8). doi:10.1145/2330601.2330605.
- Sporns, O. (2011). *Networks of the brain*. Cambridge, MA: MIT Press.
- Stevens, R., Johnson, D., & Soller, A. (2005). Probabilities and predictions: Modeling the development of scientific problem-solving skills. *Cell Biology Education*, 4(1), 42–57. doi:10.1187/cbe.04-03-0036.

- Tashakkori, A., & Teddlie. (2003). *Handbook of mixed methods in social and behavioral research*. SAGE. <http://books.google.com/books?id=F8BFOM8DCKoC&pgis=1>
- Tukey, J. W. (1977). *Exploratory data analysis. Analysis* (Vol. 2, p. 688). doi:10.1007/978-1-4419-7976-6.
- Turkay, S., & Tirthali, D. (2010). Youth leadership development in virtual worlds: A case study. *Procedia Social and Behavioral Sciences*, 2(2), 3175–3179. doi:10.1016/j.sbspro.2010.03.485.
- Van Der Pol, J., Van Den Berg, B., Admiraal, W., & Simons, P. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers and Education*, 51(4), 1804–1817. doi:10.1016/j.compedu.2008.06.001.
- Webb, M. (2010). Beginning teacher education and collaborative formative e-assessment. *Assessment and Evaluation in Higher Education*, 35(5), 597–618.
- Witten, F., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (p. 524).

Fig. 1 Screenshots of two virtual performance assessments



Fig. 2 Graphs of time on task (x) versus total score (y) in the *upper right corner*; and a selected solution on the Pareto curve (x = complexity, y = error) in the *lower right corner*. A range of solutions is presented in the *top left box* with varying degrees of complexity and error; the learning analyst selects a solution to minimize both error and complexity (the *highlighted dot* in the *lower right box*) and then later builds a network representation via a canonical process that fully details the solution as a system of nonlinear equations

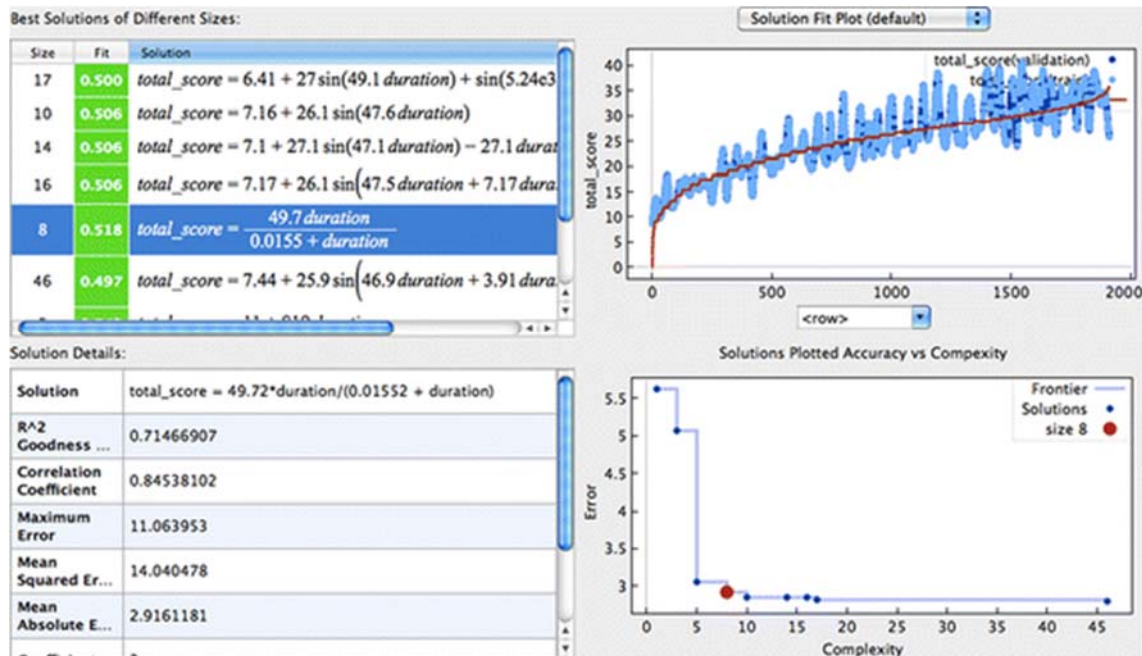


Fig. 3 Digraph of association rules in a network. Labels in this network graph represent specific activities and patterns in the digital learning experience. For example, experiment_1 might be a test of the impact of an environmental factor on the frogs (or bees). Students who conducted these specific tests in these orders reached stronger conclusions

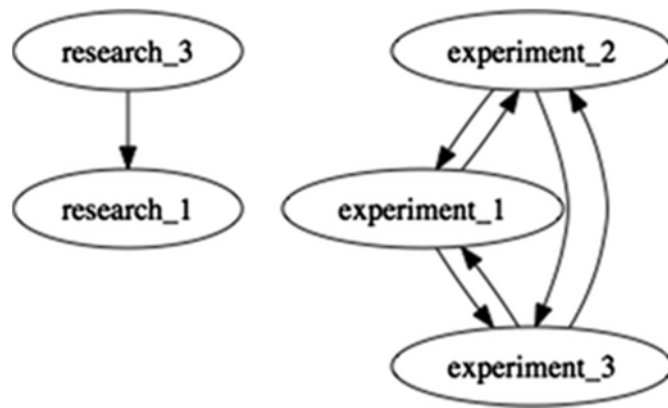


Fig. 4 Data sources, analytic data set (ADS) and self-organizing map

